Economic and Socioeconomic Determinants of Infant Mortality: A cross-country investigation

Yang Liu
Undergraduate, Coulter Department of Biomedical Engineering, Georgia Institute of Technology

Yiwen Chen
Undergraduate, Department of Mathematics, Georgia Institute of Technology

Di Wang
Undergraduate, Coulter Department of Biomedical Engineering, Georgia Institute of Technology

Abstract

Infant mortality rate has been gaining greater importance in recent years as an indicator of population wellness. In previous studies, infant mortality rate is often found to be higher in countries with lower income and less socioeconomic development. This paper concentrates on investigating possible determinants of infant mortality rate across countries. Using data of 114 countries in 2011 from the World Bank, analysis has been done to explore the relationship of infant mortality rate with economic strength and socioeconomic factors. Based on the results, in can be concluded that female education and government health expenditure are negatively related to infant mortality rate, but fertility rate is positively related to infant mortality rate. Comparing with fertility and female education, government health expenditure has less impact on infant mortality rate, especially with respect to countries with lower per capita GDP.

1. Introduction

With an increasing attention to health and disease globally, the realm of healthcare plays important roles not only on developing effective therapeutics and medicines, but also on enhancing standard of living for any nation. As improving population health being one of the ultimate goals for economic growth, a country targets on a variety of economic metrics to highlight their performance on health care other than national outcomes and production yielded. One of the widely used indicators for quality of life is the infant mortality rate (IMR), which is defined as the death of an infant before their first birthday and is an estimated number of infant deaths for every 1,000 live newborns. The infant mortality rate is one of the indicators to examine the well-being of population health as it is related to maternal health, access to medical treatments, socioeconomic conditions. Statistical analysis had displayed the fact that higher infant mortality within a population is correlated to the health within same population.

Some people agree that the apparent association between the causes of infant mortality that are likely to influence the health status of whole populations such as their economic development, general living conditions, social well-being, rates of illness, and the quality of the environment. Thus, the causes and factors affecting infant mortality that are likely to influence health status of whole populations have been investigated for comprehensive evaluations. However, using IMR as a proxy measure of population health is still a controversial topic. Recently, there are arguments about the improprieties of using IMR as a measure of population health due to the fact that it derives from a small, non-representative portion of the population and excludes any consideration of non-fatal health outcomes. As a consequence, health and other policies would begin to target on achieving goals made from this measure, while ignoring the rest of the population, for which the outcome measure is supposed to be the correct indicator. In its simplest terms, when health policies are made, the whole population’s health may still remain static or even degrade even though the IMR indicator decreases, due to the fact that lowering infant mortality becomes the principal focus of the policy.

Therefore, as more concerns regarding the infant mortality have risen in different countries, studying this specific population health indicator with most recent data statistics becomes necessary. As the situations between developed countries and developing countries are different, data are stratified and analyzed separately. The objective of this study is to investigate possible factors affecting infant mortality rates across countries, specifically female education, GDP per capita, public health expenditures and fertility rate, and to statistically show the inpact of these factors on mortality rates. We hypothesize that with higher fertility rates, lower health spending, less economic strength, and lower female education level, statistically significant increase in infant mortality rates can be observed, based on the data of 114 countries.

2. Literature Review

Previous studies indicate that infant mortality is correlated with education of women. Many countries have improved female education as one of the most “sought out Millennium Development Goals”. There have also been many policy recommendations aiming for improvements in women’s education level in order to reach lower infant mortality. Papageorgiou and Stoytcheva (2008), however, noted an interesting approach to this issue. Unlike other studies focusing on average years of education, they examined the relationship between female human capital inequality measured by the Gini coefficient (a measure of human capital inequality) and infant mortality, using a cross-country dataset and growth regression method. The article proposed the hypothesis that “higher inequality in education among women may be partly responsible for higher infant mortality”. This is because that mothers at the low end of the distribution may lack the necessary skills to provide adequate health care and appropriate living environment to their infants. The article also examined empirically whether female education inequality affects growth via infant mortality. It first tests whether inequality in women’s education leads to higher infant mortality. Then it tests whether higher infant mortality could be partly responsible for the slow growth in many developing countries. Gini coefficients and average female education for 108 countries are used for the examination. The results show a strong evidence for the first hypothesis that higher female education inequality results in higher infant mortality rate. There is also some supportive evidence for the second hypothesis that female human capital inequality could be an important obstacle to the growth in some developing countries. Both results point to a simple policy implication of an increase in human capital of the least educated women. This paper contributes to the sparse literature on how inequality in education among women account for infant mortality rate by empirically testing convergence across countries.

Many other studies have considered the relationship between economic growth and IMR. Detailed examination of this relationship by Erdoğan(2008) showed a significant negative relationship between real per capita GDP and infant mortality in selected countries. In the study, data were gathered from 25 high income OECD countries on yearly basis from 1970 to 2007, and the relationships between the variables were analyzed. Erdoğan first did panel unit root test under the assumption that data are independent to each other to test the stationarity of the variables included in the regression model in order to obtain unbiased estimations. Then, the two way fixed effects model was formulated controlling for unmeasured time-invariant differences between units and unit-invariant difference between time period. According to the empirical findings, Erdoğan found that 10% increases in per capita GDP, IMR decreases in the ratio of 28.9%, which leading to the conclusion that the infant mortality rate of the countries decrease as countries become rich and economic powers grow.

Moreover, researchers had investigated the impact of public spending on health on infant mortality rate. The effect of health expenditure had been continuously a controversial topic across countries. Some studies had claimed evidences on significant impact of public spending. However, some other studies had opposite announcement on analysis using different datasets. A study investigated the effect of government health expenditure by evaluating the cross-sectional data, extracted from World Bank for 1992 and 1993 on 98 developing countries. In the study, Filmer and Pritchett (1999) validated that the impact of public spending on health is statistically insignificant and thus is not a powerful determinant of mortality as public spending explained so little, compared to economic and social factors, which could explain 95% of a cross-national variation. It had shown that the independent variation in public spending explained only less than 1% of the differences in mortality rate. Although reverse causation and measurement error were pointed out to be potential sources of bias in estimation and two-stage least squares were done and demonstrated that attenuation bias was made due to measurement error in the OLS estimates, the results using the alternative data have still shown the same results. Thus, Filmer and Pritchett conclude that a small impact of government health spending would translate into low explanatory power to the mortality outcome. In addition, the possible factors they identified are mainly income per capita, income inequality, female education, ethnic fractionalization, and religions.

Zakir and Wunnava (1999) tested causality between infant mortality rates and fertility rates, but demonstrates that fertility rates do have for factors affecting infant mortality rates based on a cross-sectional model covering 117 countries for the year 1993. They examined the effects of fertility rates, female participation in the labor force, per capita GNP, female literacy rates, and government expenditure on health-care, as a percentage of GNP, on infant mortality rates across countries. According to the generalized least squares regression results, all other factors significantly affect infant mortality rates with the exception of expenditure on health programs. Furthermore, infant mortality rates are strongly negatively correlated with female literacy rates. Also, they demonstrate that their findings contradict with a previous argument stating that there is a dual a big effect on infant mortality rates.

It appears that the factors affecting infant mortality rates might change over time, as the society evolves and the technologies are developed quickly. Thus, with limited access to global data of education, public health expenditure, GDP per capita, and fertility rate, the results of previous literatures might had been outdated. Thus, rendering more statistical analysis worldwidely for recent years, this paper is aimed to test these factors in a more accurate manner, in order to validate the results, and correct possible biased conclusions from previous studies from a global perspective.

3. Data and Methodology

In order to effectively investigate the relationship between IMR and suspected responsible socioeconomic factors, we analyzed data by using univariate and multivariate regression. We first run regression on data that includes all countries used in this study, and then run the other regression on data stratified for income levels of countries. In the regression models, IMR will serve as the predicted variable. Four predictors are used in the study: total fertility rate (birth per woman), government health expenditure (as a percentage of total spending) and maternal education (female gross enrollment ratio of the tertiary education as socioeconomic factors, and GDP per capita (in 2005 USD) as economic strength measure. Table 1 is a summary of the variables.

Table 1 Variables and Descriptions

Variables	*Description*
IMR: infant mortality rate	Number of deaths in infant aged under one year per 1,000 live births.
fert: total fertility rate	Births per women. Calculated by dividing the number of births by number of female population.
hexp: health expenditure	Calculated as a percentage of total government spending rather than dollars. This is for the consideration of countries' absolute consumption differences due to different income level
gdppc: GDP per capita	Annual percentage growth rate of GDP per capita based on constant 2005 U.S. dollars.
educ	Female gross enrollment ratio of the tertiary education.

A strong positive linear association can be found between *IMR* and *fert* (Fig.1). However, there is a nonlinear relationship between *IMR* and *educ*. Thus we take log-transform of *educ* and estimate its relationship with *IMR* on original scale. The association between predicted *IMR* and log-transformed *educ* predictor appears to be strongly and negatively linear (Fig.2). The relationship between *IMR* and *hexp* is found to be negatively linear (Fig.3). Even though this association is relatively weak, it is significant at 95% level. However, the linear association between *IMR* and *gdppc* is not significant at either 95% (Fig.4) or 90% level.

The following equations are estimated for the purpose of exploring relationships between dependent variable and each independent variable.

$$IMR = \alpha_1 + \beta_1 fert + \mu_1 \ (1)$$
$$\text{IMR} = \alpha_2 + \beta_2 hexp + \mu_2 \ (2)$$

$$\text{IMR} = \alpha_3 + \beta_3 gdppc + \mu_3 \; (3)$$
$$\text{IMR} = \alpha_4 + \beta_4 \ln(educ) + \mu_4 \; (4)$$

Figure 1 Relationship between IMR and fertility rate

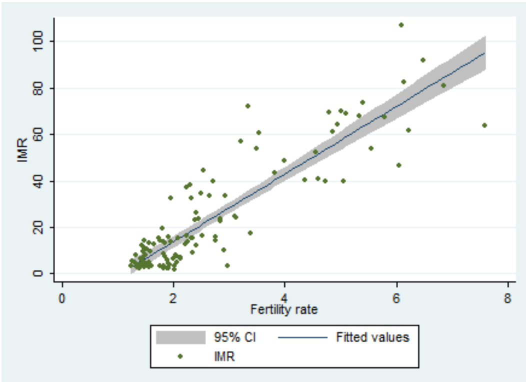


Figure 2 Relationship between IMR and ln(educ)

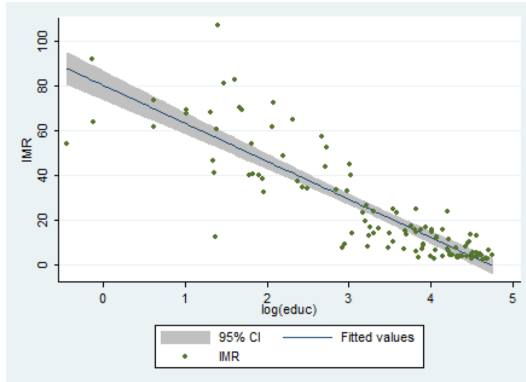


Figure 3 Relationship between IMR and hexp

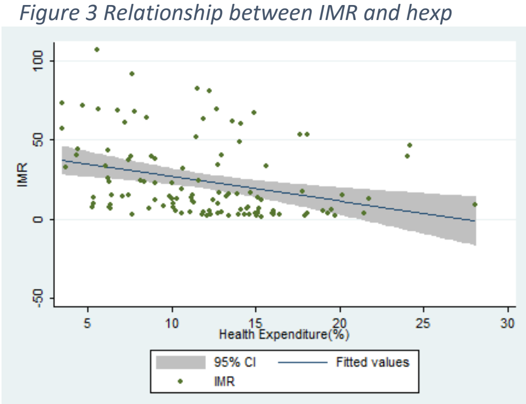
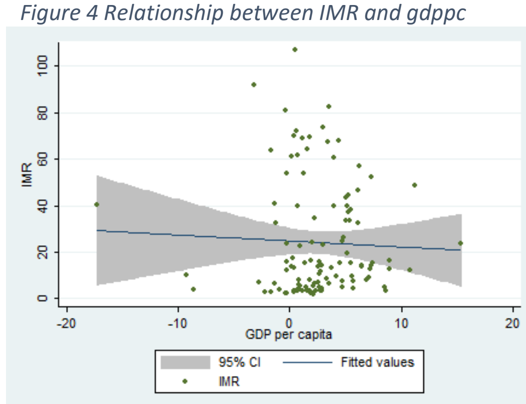


Figure 4 Relationship between IMR and gdppc



To detect outliers, DFBETA (Standardized difference of betas) with cutoff of 1.00 is used for all repressors. Only one outlier has been detected. Figure 5 shows the plot of DFBETA. The leverage versus the squared residuals is also provided (Fig.6). As we can see, a couple of countries have either high leverage or large residuals.

Figure 5 DFBETA

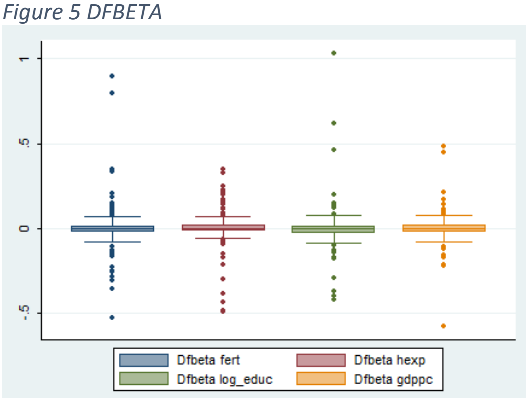
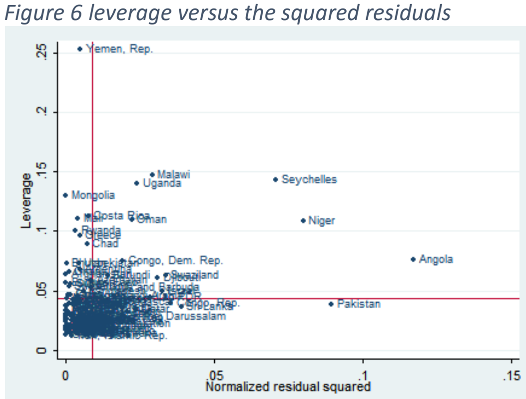


Figure 6 leverage versus the squared residuals



The following equations are estimated for investigating correlations between IMR and socioeconomic factors as adding more predictors.

$$\text{IMR} = \alpha_0 + \beta_f fert + \beta_h hexp + u \text{ (I)}$$

$$\text{IMR} = \alpha_0 + \beta_f fert + \beta_h hexp + \beta_e \ln(educ) + u \text{ (II)}$$

In order to better understand the relationship between IMR and GDP per capita as economic strength factor, we first hypothesized OLS model on all of selected countries (Equation. III), and then run regressions on data of countries stratified for lower and higher GDP per capita (Equation IV&III). The hypothesized OLS models are as follows:

$$\text{IMR} = \alpha_0 + \beta_f fert + \beta_h hexp + \beta_e \ln(educ) + \beta_g gdppc + u \text{ (III)}$$

$$IMR_1 = \alpha_{0,1} + \beta_{f,1} fert + \beta_{h,1} hexp + \beta_{e,1} \ln(educ) + u_1 \text{ (IV)}$$

$$IMR_2 = \alpha_{0,2} + \beta_{f,2} fert + \beta_{h,2} hexp + \beta_{e,2} \ln(educ) + u_2 \text{ (V)}$$

Data are gathered from the World Bank database of year 2011. A total of 114 countries and regions are selected for analysis. Other countries and regions are excluded due to incomplete and missing values of the predictors being used in this study. Table 2 presents the descriptive statistics of the variables.

Table 2 Descriptive statistics for selected variables

Variables	*N*	*Mean*	*Std Dev*	*Min*	*Max*
IMR	114	24.17456	24.39886	1.8	106.8
fert	114	2.702013	1.496094	1.244	7.581
hexp	114	11.94715	4.82793	3.471622	28.06736
gdppc	114	2.575939	3.970595	-17.34115	15.31755
ln(educ)	114	3.281918	1.249605	-.4459278	4.753324

4. Results

4.1 Empirical Estimates

Table 3 shows the results of simple regressions (Equation1-4). The R-squared values of 0.0952 and 0.0017 indicate weak relationships between *IMR* and corresponding predictors. The coefficient of *gdppc* is not significant, and the other coefficients are all significant at 1% level. The results for simple regression can be interpreted as follows: For each additional infant birth per women, *IMR* increases by 14.528; Each additional percentage point in government health expenditure reduces *IMR* by -1.5590; one percent increase in female gross enrollment ratio of the tertiary education results in 0.169998 percent decrease in *IMR*.

Table 3. Parameter estimates for simple regressions on all data

Model #	Regressor	df	Coef.	_cons	Std. Err.	t	P>|t|	R-squared
1	fert	113	14.528	-15.0803	.7001	20.75	0.000***	0.7936
2	hexp	113	-1.5590	42.8002	.4542	-3.43	0.001***	0.0952
3	gdppc	113	-.2570	24.8366	.5801	-0.44	0.659	0.0017
4	ln(educ)	113	-16.9998	79.9664	.9075	-18.73	0.000***	0.7580

*Statistical significance denoted using asterisks: *P<0.10, **P<0.05, ***P<0.01*.

Multiple regressions are run to explore the interaction among variables as more predictors are added (Equation I-V). This first regression run is OLS regression with all countries selected. The results are listed in Table 4-8. The R^2 values indicate moderately strong associations between *IMR* and the independent variables in these models.

In Model I (Table 4), we add *hexp*. The results show that one percent increase in government health expenditure leads to reduction of -0.9136 in *IMR*. The estimates are significant at 1% level. The magnitude of coefficient of *hexp* decreases from 1.5590 to 0.9136.

Table 4. Parameter estimates for regression model I

Regressor	Coef.	Std. Err.	t	P>|t|	R-squared = 0.8255 Adj R-squared = 0.8224 Df = 113
fert	14.0923	.6537	21.56	0.000***	
hexp	-.9136	.2026	-4.51	0.000***	
_cons	-2.9883	3.3419	-0.89	0.373	

*Statistical significance denoted using asterisks: *P<0.10, **P<0.05, ***P<0.01*.

In Model II (Table 5), we add *ln(educ)*. The estimates on regressors all remain their signs and significance. The magnitude of *fert* decreases from 14.0923 to 9.5272, and the magnitude of *ln(educ)* decreases from 16.9998 to 6.4249. The result implies: a unit increase of *fert, hexp, and ln(educ)* leads to 9.5272, -0.7189 and -6.4249 percent change in *IMR*, respectively.

Table 5. Parameter estimates for regression model II

Regressor	Coef.	Std. Err.	t	P>|t|	R-squared = 0.8508 Adj R-squared = 0.8467 Df = 113
fert	9.5272	1.2198	7.81	0.000***	
hexp	-.7189	.1935	-3.72	0.000***	
ln(educ)	-6.4249	1.4889	-4.31	0.000***	
_cons	28.1076	7.8468	3.58	0.001***	

*Statistical significance denoted using asterisks: *P<0.10, **P<0.05, ***P<0.01*.

In Model III (Table 6), *fert* is omitted. The estimates of *hexp* and *ln(educ)* remain negative and significant. The effect size of *ln(educ)* increases greatly by omitting *fert*. But R^2 decreases to 0.7681. In Model IV (Table 7), *hexp* is excluded while *fert* and *ln(educ)* are retained. The estimates remain their original signs and significance. Generally, the magnitude of *ln(educ)* and *fert* decreases when they are both included.

Table 6. Parameter estimates for regression model III

Regressor	Coef.	Std. Err.	t	P>|t|	R-squared = 0.7681 Adj R-squared = 0.7639 Df = 113
hexp	-.5216	.2381	-2.19	0.031**	
ln(educ)	-16.5105	.9200	-17.95	0.000***	
_cons	84.5928	3.7781	22.39	0.000***	

*Statistical significance denoted using asterisks: *P<0.10, **P<0.05, ***P<0.01*.

Table 7. Parameter estimates for regression model IV

Regressor	Coef.	Std. Err.	t	P>|t|	R-squared = 0.8321 Adj R-squared = 0.8291 Df = 113
fert	8.9355	1.2773	7.00	0.000***	
ln(educ)	-7.7142	1.5292	-5.04	0.000***	
_cons	25.3482	8.2499	3.07	0.003***	

*Statistical significance denoted using asterisks: *P<0.10, **P<0.05, ***P<0.01*.

Table 8 shows the results of Model V which includes all variables for all countries selected. The R^2 value of 0.8547 indicates a moderately strong fit of data. The estimates of *fert*, *hexp and ln(educ)* continue to be negative and significant at 1% level. However, the estimate of *gdppc* becomes positive even though the magnitude remains small. A summary of results of Model I-V are presented in Table 9.

Table 8. Parameter estimates for regression model V

Regressor	Coef.	Std. Err.	t	P>|t|	R-squared = 0.8547 Adj R-squared = 0.8494 Df = 113
fert	9.7908	1.2190	8.03	0.000***	
hexp	-.7262	.19188	-3.78	0.000***	
ln(educ)	-6.2392	1.4800	-4.22	0.000***	
gdppc	.3887	.2268	1.71	0.089*	
_cons	25.8716	7.8874	3.28	0.001***	

*Statistical significance denoted using asterisks: *P<0.10, **P<0.05, ***P<0.01*.

Table 9. Parameter estimates for OLS regression on all countries selected

Dependent variable: IMR

Model	(I)	(II)	(III)	(IV)	(V)
Constant	-2.9883*** (3.3419)	28.1076*** (7.8468)	84.5928*** (3.7781)	25.3482*** (8.2499)	25.8716*** (7.8874)
fert	14.0923*** (.6537)	9.5272*** (1.2198)		8.9355*** (1.2773)	9.7908*** (1.2190)
hexp	-.9136*** (.2026)	-.7189*** (.1935)	-.5216** (.2381)		-.7262*** (.19188)
ln(educ)		-6.4249*** (1.4889)	-16.5105*** (.9200)	-7.7142*** (1.5292)	-6.2392*** (1.4800)
gdppc					.3887* (.2268)
Adj. R^2	0.8224	0.8467	0.7639	0.8291	0.8494
Obs.	114	114	114	114	114

*Standard errors are in parentheses. Statistical significance denoted using asterisks: *P<0.10, **P<0.05, ***P<0.01.*

To further explore the impact of *gdppc* on *IMR*, separate regressions are run using data stratified for countries with lower and higher GPD per capita separately. The data are divided into halves according to *gdppc*. The OLS regression was run with independent variables *fert*, *hexp and ln(educ)* on 50% countries with higher *gdppc* and 50% countries with lower *gdppc* separately. The results are summarized in Table 10.

Table 10. Parameter estimates for OLS regression on stratified data according to gdppc

Dependent variable:	**Specification Estimation**	**Constant**	**fert**	**hexp**	**ln(educ)**	**Adj. R^2**	**Obs.**
IMR	(VI)	26.3559* (14.0000)	9.7645*** (2.0573)	-.8582** (.3339)	-5.9154** (2.5052)	0.8516	57
	(VII)	28.0453*** (9.0389)	9.6317*** (1.5160)	-.5835** (.2312)	-6.6120*** (1.7752)	0.8341	57

*Standard errors are in parentheses. Statistical significance denoted using asterisks: *P<0.10, **P<0.05, ***P<0.01.*

The regression on countries with lower *gdppc* has a slightly higher R^2 value than the regression on countries with higher gdppc. The estimates of *hexp* are significant at 5% level and the estimates of *fert* are significant at 1% for both models. The estimates of *ln(educ)* are significant at 5% level for specification (VI) and 1% level for specification (VII). The magnitudes of predictors of specification (VII) are slightly greater than specification (VI).

4.2 Robustness analysis

We begin robustness analysis by doing some diagnosis on the OLS regression. The Cook's distance test is used with cutoff of 4/114. The countries with Cook's D values larger than 1 are excluded in the corresponding robust regression analysis. The results of robust regression are showed in Table 11. Comparing with OLS regression, the results slightly different. Particularly, the robust estimate of *hexp* is non-significant for Specification III and VI. In summary, the OLS results showed in Table 9&10 are fairly robust.

Table 12. Robust regression

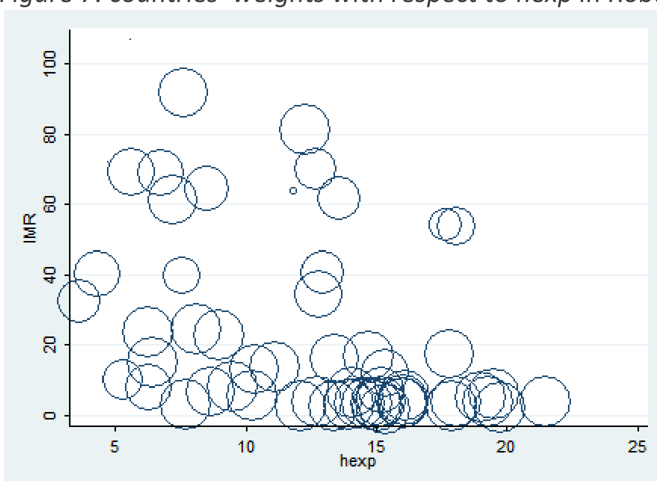
Dependent variable:	**Specification Estimation**	**Constant**	**fert**	**hexp**	**ln(educ)**	**gdppc**	**F**	**Prob>F**	**Obs.**
IMR	(I)	-5.2524* (2.9490)	14.333*** (.5768)	-.8140*** (.1787)			F(2,111) = 343.26	0.0000***	114
	(II)	36.5654*** (5.9430)	8.6098*** (.9239)	-.3238** (.1466)	-9.3732*** (1.1277)		F(3,110) = 419.07	0.0000***	114
	(III)	80.4092*** (2.9082)		-.2976 (.1833)	-16.2656*** (.7082)		F(2,111) = 291.28	0.0000***	114
	(IV)	35.1237*** (5.7815)	8.6511*** (.8951)		-10.1095*** (1.0717)		F(2,111) = 690.65	0.0000***	114

(V)	33.4116*** (5.7112)	8.8756*** (.8826)	-.2457* (.1389)	-9.3828*** (.1389)	.5692*** (.1642)	F(4,109) = 354.59	0.0000***	114
(VI)	20.4354* (11.2160)	10.1620*** (1.6482)	-.3634 (.2675)	-6.4564*** (2.007)		F(3,53) = 172.68	0.0000***	57
(VII)	33.1791*** (6.3517)	9.1033*** (1.0653)	-.2831* (.1624)	-8.3792*** (1.2474)		F(3,53) = 212.77	0.0000***	57

*Standard errors are in parentheses. Statistical significance denoted using asterisks: *P<0.10, **P<0.05, ***P<0.01.*

To better visualize the countries' weights in the robust regression with respect to the coefficients of *hexp* in specification VI, the graph of data points with the weight information as the size of circles is provided in Figure 7. A smaller circle indicates the country is down-weighted in the robust regression. In this case, Angola, Pakistan and Niger are three countries being most down-weighted.

Figure 7. countries' weights with respect to hexp in Robust(VI)



Note: the size of circles indicates the magnitude of weights

5. Conclusion

Infant mortality rate is an important indicator of the general wellness of a country. In this paper, we investigate the relationships of IMR with economic factor as GDP per capita and socioeconomic factors as fertility rate, female education, and government health expenditure. We start by investigating univariate regressions and then run multivariate regressions on all the data gathered. To further explore the impact of per capita GDP on IMR, the data are stratified based on per capita GDP and analyzed using OLS models.

The results obtained have suggested that IMR is positively associated with fertility, and negatively associated with government health expenditure and female education. In the case of countries with higher GDP per capita, fertility rate, government health expenditure and maternal education are significant determinants of IMR. Whereas in the case of countries with lower GDP per capita, only fertility rate and maternal education are significant indicators of IMR. It is interesting that the sign of GDP per capita is positive unexpectedly. After delving deeper into literature, we found that the distribution of wealth within each country can also affects IMR, not simply the overall wealth of countries. For instance, the US has the highest IMR in the OECD countries due to economic inequality. Results further indicate that fertility rate and female education have stronger impacts on IMR than government health expenditure and per capita GDP. Future research should be done to better understand the impact of economic strength on IMR. For example, GNI per capita, which has recently been defined as income classification standards of countries by the World Bank, can be used as the economic indicator in future studies.

Acknowledgement

We would like to extend our thanks to Dr Shatakshee Dhongde for her help and support in completing the paper.

References

Erdoğan E, Ener M, Arıca F. (2013). The Strategic Role of Infant Mortality in the Process of Economic Growth: An Application for High Income OECD Countries. Procedia - Social and Behavioral Sciences. Volume 99, 6 November 2013, Pages 19–25

Filmer D, Pritchett L. 1999. The impact of public spending on health: does money matter? Social Science and Medicine 49(10): 1309–1323.

Papageorgiou, C. and Stoytcheva, P. (2008). Education Inequality among Women and Infant Mortality: A cross- country empirical investigation. Available at http://www.chrispapageorgiou.com/papers/Female-Mortality.pdf (26.01.2011).

Zakir, Mohammed, and Phanindra V. Wunnava. "Factors Affecting Infant Mortality Rates: Evidence from Cross–sectional Data." Applied Economics Letters 6.5 (1999): 271-73. Web.